BAKER BOTTS L.L.P.

30 ROCKEFELLER PLAZA

NEW YORK, NEW YORK 10112

TO ALL WHOM IT MAY CONCERN:

Be it known that WE, TERRY THOMAS, MICHAEL NUCCIO and TZUNG-FU HSIEH, citizens of the United States of America; the United States of America; and the People's Republic of China; respectively, residing in College Station, County of Brazos, State of Texas; Durham, County of Durham, State of North Carolina; and Burlingame, County of San Mateo, State of California; respectively, whose post office addresses are 2804 Cloister Drive, College Station, Texas, 77845; 1022 Edinborough Drive, Durham, North Carolina, 27703; and 1471 El Camino Real, Burlingame, California, 94010, respectively, have invented an improvement in

CONSTITUTIVE PROMOTER FROM *ARABIDOPSIS*

of which the following is a

SPECIFICATION

BACKGROUND OF THE INVENTION

[0001] The present invention is directed to a promoter derived from an *Arabidopsis* gene. The promoter of the invention is useful in transgenic organisms in which a high level of production of a gene product is desired.

[0002] Promoters are regulatory elements that direct the expression of genes. Both constitutive and regulated promoters are used to direct gene expression in transgenic organisms including plants. Constitutive promoters direct expression in most or all tissues, and are useful when high levels of production of a gene product are desired. The 35S promoter from cauliflower mosaic virus (CMV) is frequently used to direct constitutive expression. Regulated promoters, such as tissue-specific and inducible promoters, are used to direct spatially or temporally specific expression, or expression in response to environmental factors.

[0003] The present invention is directed to a promoter that directs constitutive expression of genes in plants. The present promoter is derived from a gene of *Arabidopsis* thaliana. In accordance with the present invention, it has been discovered the subject promoter can direct high levels of constitutive expression of heterologous genes in plants.

[0004] The present invention is directed to an isolated promoter derived from a gene of *Arabidopsis* thaliana that encodes an endomembrane associated (ENDO) gene. In a preferred embodiment, the promoter has at least 70% identity to the sequence of SEQ ID NO:1. The present invention further provides a nucleic acid construct comprising the promoter of the invention operably linked to a heterologous nucleic acid. Vectors comprising the nucleic acid construct are also provided. In another embodiment, the present invention is directed to a plant cell comprising the nucleic acid construct of the invention. Transgenic plants and progeny thereof comprising the construct, parts of such plants, and methods of making such plants, are also provided.

DETAILED DESCRIPTION OF THE INVENTION

[0005] The present invention is directed to a promoter that directs expression of genes in plants.

[0006] The promoter is derivable from an *Arabidopsis* gene (ENDO) that encodes an endomembrane-associated protein. In accordance with the present invention, the promoter is derivable from the ENDO gene and is designated the ENDO promoter.

[0007] In a preferred embodiment, the ENDO promoter has the sequence of SEQ ID NO:1, or a fragment thereof that has promoter activity, i.e., drives the expression of a heterologous nucleic acid operable linked thereto.

[0008] In another preferred embodiment, the ENDO promoter has a sequence that has at least 70% identity to the sequence of SEQ ID NO:1 or a fragment thereof that has promoter activity. More preferably, the ENDO promoter has a sequence that has at least 80%, or more preferably at least 90%, identity to the sequence of SEQ ID NO:1 or a fragment thereof that has promoter activity. Sequence identity as defined herein is measured using the program Clustal W described by Thompson et al. (1994) Nucleic Acid Research 22:4673 and may be calculated using the EMBL Nucleotide Sequence Database (http://www.ebi.ac.uk/embl.html).

[0009] In another embodiment, the ENDO promoter has a nucleic acid sequence that hybridizes to the sequence of SEQ ID NO:1 under high stringency conditions and that has promoter activity. High stringency conditions are defined herein as 68°C in buffered aqueous solution or 42°C in 50% formamide.

[0010] The promoters of the present invention may be isolated by using a nucleic acid having the sequence of SEQ ID NO:1 or a fragment thereof to probe a plant genomic library. In a preferred embodiment, the library is an *Arabidopsis* genomic library. Such libraries may be made by well-known methods disclosed for example in Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. or obtained as a bacterial artificial chromosome (BAC) genomic library from the Arabidopsis Biological Resource Center (ABRC) at Ohio State University, 173S Neil Avenue, Columbus, Ohio 43210. The probes can be used to isolate nucleic acids that hybridize to SEQ ID NO:1 under high stringency conditions. High stringency conditions are described in Sambrook et al., id., and Beltz et al. (1983) Methods Enzymol. 100:226 and include, for example, hybridization at 68°C in aqueous buffered solution or at 42°C in 50% formamide. Having

identified a genomic clone, the promoter can be derived by endonuclease or exonuclease digestion, or PCR amplification.

[0011] Further, probes derived from SEQ ID NO:1 may be used to isolate promoters having at least 70%, or at least 80%, or at least 90% identity to SEQ ID NO:1. In addition, well-known methods of enzymatic and chemical synthesis and modification of nucleic acids may be used to obtain promoters having the stated levels of identity to SEQ ID NO:1 or fragments thereof.

[0012] The present invention is also directed to methods for isolating a promoter derived from the *Arabidopsis thaliana* ENDO gene comprising (a) probing a plant genome with a nucleic acid having the sequence of SEQ ID NO:1, (b) hybridizing said nucleic acid to a nucleic acid of the plant genome under high stringency conditions, and (c) isolating the promoter from the plant genome. Preferably, the plant genome is an *Arabidopsis* genomic library.

[0013] The present invention also encompasses fragments of SEQ ID NO:1 and sequences having at least 70% identity thereto that have promoter activity. Those of ordinary skill in the art can determine the sequence required to maintain promoter activity, for example by generating deletion fragments of SEQ ID NO:1 to obtain putative promoters, operable fusing the putative promoter to a transgene, introducing the construct into a host cell, and measuring expression of the transgene. The transgene may be a reporter, for example, the chloramphenicol acetyl transferase (cat), beta-glucuronidase (gus) or luciferase (luc) genes. The construct containing the promoter and transgene is cloned into a vector, and the vector is used to transform host cells. Expression of the transgene is measured by assaying for the transgene product. Standard assays are available to sensitively detect the reporter gene product. For example, GUS can be measured by histochemical or fluorogenic assays. Jefferson et al. (1987) EMBO J. 6:3901. The presence of the transgene product is indicative of a functional promoter.

[0014] The present invention further provides a nucleic acid construct comprising the promoter operably linked to a heterologous nucleic acid. The heterologous nucleic acid is any nucleic acid other than the ENDO gene. As used herein, the term heterologous nucleic acid includes all synthetically engineered and biologically derived genes which may be introduced into a plant by genetic engineering, including but not limited to nonplant genes, plant genes, modified genes, synthetic genes and portion of genes. The heterologous nucleic acid preferably contains the coding region of a protein or polypeptide or antisense molecule of interest. Suitable heterologous nucleic acids for use herein include all nucleic acids that will provide or enhance a beneficial feature of the resultant transgenic plant. For example, the nucleic acid may encode proteins or antisense RNA transcripts in order to promote increased food values, higher yields, pest resistance, disease resistance, herbicide tolerance, and the like. Representative nucleic acids include, for example, a bacterial dap A gene for increased lysine; genes that encode *Bacillus thuringiensis (Bt)* endotoxins (inter alia U.S. Patent Nos. 5,460,963; 5,683,691; 5,545,565; 5,530,197; 5,317,096) or insecticidal toxins isolated from *Photorhabdus* (WO97/17432 or WO98/08932) for insect resistance; lytic peptides genes for disease resistance, genes imparting tolerance to oxynil herbicides (U.S. Patent Nos. 4,810,648 and 5,559,024), bacterial or plant EPSPS for resistance to glyphosate and EPSPS inhibitor herbicides (U.S. Pat. Nos. 4,940,835; 5,188,642; 4,971,908; 5,145,783; 5,312,910; 5,633,435; 5,627,061; 5,310,667, WO 97/04103); genes imparting tolerance to glufosinate (EP 242 236) bacterial or plant HPPD (WO 96/38567, WO 98/02562) for resistance to HPPD-inhibitor herbicides (i.e. diketones, isoxazoles, etc.), chitinase or glucan endo 1, 3-B-glucosidase for fungicidal properties. Also, the nucleic acid may be introduced to act as a genetic tool to generate mutants and/or assist in the identification, genetic tagging, or isolation of segments of monocot genes.

[0015] As a preferred embodiment of the present invention, the heterologous nucleic acid encodes a protein to impart herbicide tolerance, more preferably tolerance to an oxynil herbicide (disclosed in U.S. Patent Nos. 4,810,648 and 5,559,024), to EPSPS inhibitor herbicides, including glyphosate and its various salts (disclosed in U.S. Patent Nos. 4,535,060; 4,769,061; 5,094,945; 4,940,835; 5,188,642; 4,971,908; 5,145,783; 5,312,910; 5,310,667; 5,633,435; 5,627,061; 5,554,798; 5,633,448; WO 97/04103), to glufosinate (EP 242 236), or to HPPD inhibitors (WO 96/38567 and WO 98/02562). More preferably, the heterologous nucleic acid encodes a protein to impart tolerance to EPSPS inhibitor herbicides.

[0016] As another preferred embodiment of the present invention, the heterologous nucleic acid 2o encodes a protein to impart insect resistance, more preferably genes which encode for *Bacillus thuringiensis (Bt)* endotoxins (inter alia, U.S. Patent Nos. 5,460,963; 5,683,691; 5,545,565; 5,530,197; 5,317,096). The nucleic acids that are preferably embraced by the instant invention are *cryl, cryll, cryIll,* and *cryIV* genes. More preferably, the genes include: *cryIA(a), cryIA(b), cryIA(c)*; and *cryIII(a)*. Most preferably the gene is *cryIA(a), cryIA(b) or cryIA(c)*. The nucleic acid construct comprising the promoter operably linked to a heterologous nucleic acid may be constructed by methods well-known in the art. The term "operably linked" as used herein means that the promoter and heterologous nucleic acid are oriented such that the promoter directs expression of the heterologous nucleic acid, generally in the 5'- to 3'-direction. The constructs may also contain polyadenylation sites at the 3'- end of the heterologous gene.

[0017] In another embodiment, the present invention provides vectors comprising the promoters and nucleic acid constructs of the present invention. The vectors may be derived from plasmids, cosmids, bacteriophage and viruses. The vectors include direct DNA delivery vectors, and vectors for *Agrobacterium*-mediated gene transfer. Direct DNA delivery vectors and

Agrobacterium based vectors, and methods for their construction, are well-known in the art and disclosed for example in "Gene Transfer to Plants", Potrykus et al., eds., Springer-Verlag, Berlin 1995 and "Plant Molecular Biology: A Practical Approach", Shaw, ed., IRL Press, Oxford 1988.

[0018] Vectors for direct DNA delivery generally contain the nucleic acid construct of the invention in a selectable bacterial replicon, and may further contain additional regulatory elements, reporter genes, and selectable markers. Vectors for *Agrobacterium*-mediated gene transfer generally contain functions to allow maintenance in *E. coli* and *Agrobacterium*, transfer from *E. coli* to *Agrobacterium*, and, *Agrobacterium* T-DNA border fragments. The vectors may be integrative or binary vectors. In a preferred embodiment, the vector is a binary vector for *Agrobacterium*-mediated gene transfer.

[0019] The vectors may further contain selectable markers and reporter genes to facilitate identification and selection of transformed cells, and suitable regulatory sequences to enable expression in plants. Weising et al. (1988) Annual Rev. Genetics 22:241 describe components that may be included in the subject vectors such as polyadenylation sequences, marker genes, reporter genes, enhancers, and introns.

[0020] The present vectors will generally contain either a selectable marker or a reporter gene or both to facilitate identification and selection of transformed cells. Alternatively, the selectable marker may be carried on a separate vector and used in a co-transformation procedure. Both selectable markers and reporter genes may be flanked with appropriate regulatory sequences to enable expression in plants. Useful selectable markers are well known in the art and include, for example, antibiotic and herbicide resistance genes. Specific examples of such genes are disclosed in Weising et al., supra. A preferred selectable marker gene is the hygromycin B phosphotransferase (hpt) coding sequence, which may be derived from *E. coli*. Other selectable

markers known in the art include aminoglyvoside phosphotransferase gene of transposon Tn5 (AphII) which encodes resistance to the antibiotics kanamycin, neomycin, and G418, as well as those genes which code for resistance or tolerance to glyphosate, bialaphos, methotrexate, imidazolinones, sulfonylureas, bromoxynil, dalapon, and the like. Selectable marker genes that confer herbicide tolerance are also of commercial utility in the resulting transformed plants.

[0021] To determine whether a particular combination of heterologous nucleic acid and recipient plant cells are suitable for use herein, the vector may include a reporter gene. Reporter genes which encode easily assayable marker proteins are well known in the art. In general, a reporter gene is a gene which is not present or expressed by the recipient organism or tissue and which encodes a protein whose expression is manifested by some easily detectable property, e.g., phenotypic change of enzymatic activity. Examples of such genes are provided in Weising et al., supra. Preferred genes include the chloramphenicol acetyl transferase (cat) gene from Tn9 of *E. coli*, the beta-gluronidase (gus) gene of the uidA locus of *E. coli*, the green fluorescence protein (GFP) gene from *Aequoria victoria*, and the luciferase (luc) gene from the firefly *Photinus pyralis*. An assay for expression of the reporter gene may be performed at a suitable time after the heterologous nucleic acid has been introduced into the recipient cells. A preferred such assay entails the use of the *E. coli* beta-glucuronidase (gus) gene described by Jefferson et al. (1987) EMBO J. 6:3901, incorporated herein by reference. Other elements such as introns, enhancers, polyadenylation sequences and the like, may also be present in the nucleic acid. These elements must be compatible with the remainder of the gene constructions. Such elements may or may not be necessary for the function of the gene, although they may provide a better expression or functioning of the gene by effecting transcription, stability of the mRNA, or the like. Such elements may be included in the nucleic acid as desired to obtain the optimal

performance of the transforming gene in the plant. For example, the maize Adhl S first intron may be placed between the promoter and the coding sequence of a particular heterologous nucleic acid. This intron, when included in a gene construction, is known to generally increase expression in maize cells of a protein. (Callis et al. (1987) Genes Dev. 1:1183). Other suitable introns include the first intron of the shrunken-1 gene of maize (Maas et al. (1991) Plant Mol. Biol. 16:199); the first intron of the castor bean catalase (cat-1) gene (Ohta et al. (1990) Plant Cell Physiol. 31:805); potato catalase second intron of the ST-LSI gene (Vancanneyt et al. (1990) Mol. Gen. Genet. 220:245); tobacco yellow dwarf virus DSV intron (Morris et al. (1992) Virology 187:633; actin-1 (act-1) intron from rice (McElroy et al. (1990) Plant Cell 2:163); and triose phosphate isomerase (TPI) intron 1 (Snowden et al. (1996) Plant Mol. Biol. 31:689). However, sufficient expression for a selectable marker to perform satisfactorily can often be obtained without an intron. (Battraw et al. (1990) Plant Mol. Biol. 15:527).

[0022] Transcription activators such as enhancers include the tobacco mosaic virus (TMV) translation activator (WO87/07644) and the tobacco etch virus (TEV) translation activator (Carrington et al. (1990) J. Virol. 64:1590). Polyadenylation and terminator regulation sequences include sequences of bacterial origin, such as the nopaline synthase (nos) terminator of *Agrobacterium tumifaciens*, or of plant origin such as the histone terminator (EP 0633317). The vector comprising the heterologous nucleic acid may also comprise sequences coding for a transit peptide, to drive the protein coded by the heterologous gene into the chromoplasts of the plant cells. Such transit peptides are well known to those of ordinary skill in the art, and may include single transit peptides, as well as multiple transit peptides obtained by the combination of sequences coding for at least two transit peptides. One preferred transit peptide is the Optimized Transit Peptide disclosed in U.S. Pat. No. 5,635,618, comprising in the direction of transcription

a first DNA sequence encoding a first chloroplast transit peptide, a second DNA sequence encoding an N-terminal domain of a mature protein naturally driven into the chromoplasts, and a third DNA sequence encoding a second chloroplast transit peptide. The constructs of the present invention are introduced into plant cells by methods known in the art. Direct gene transfer methods include gene transfer to protoplasts by microinjection, electroporation, chemically-induced DNA uptake (Potrykus, supra) and biolistic (microprojectible bombardment) approaches (Klein et al. (1987) Nature 327:70). *Agrobacterium* mediated gene transfer methods include leaf disk transformation, protoplast culture, transformation of seed, stem or root explants, *in planta* vacuum-infiltration (Potrykus, supra), and transformation of inflorescence (U.S. Patent No. 6,037,522).

[0023] Plant cells into which the nucleic acids of the present invention include cells of all plants into which nucleic acids can be transferred. Plant cells include undifferentiated tissues such as calli and differentiated tissues such as embryos, plant portions, plants and seeds. Monocotyledous and dicotyledonous plants are included. In a preferred embodiment the plant is cotton, rice, corn, wheat, barley, oat, rye, oil seed rape, potato, soybean, sunflower, sugar cane, sugar beet, alfalfa, or banana. In a more preferred embodiment, the plant is cotton, corn, or potato, and most preferably cotton.

[0024] The promoters, nucleic acid constructs, vectors, and plant cells of the present invention are useful for making recombinant gene products *in vitro*, and for making transgenic plants with desirable properties.

[0025] Another aspect of the invention provides transgenic plants, progeny thereof, and seeds and other parts thereof containing the nucleic acid construct of the present invention. Both monotyledous and dicotyledonous plants are contemplated. Plant cells are transformed with the

nucleic acid construct by any of the plant transformation methods described above, and regenerated into a complete transgenic plant by methods well known to those of ordinary skill in the art (Potrykus, supra, Shaw, supra). For *in planta* transformation methods, the regeneration step is not needed. Generally, germinating seeds or wounded plants are inoculated with *Agrobacterium* containing the nucleic acid construct, plants are grown to maturity, and seeds are collected, sown, and transformants are selected.

[0026] A method of making a transgenic plant comprising the nucleic acid construct of the present invention comprises transforming a plant cell with a vector comprising the ENDO promoter operably linked to a heterologous gene to provide a transformed plant cell, and regenerating a transgenic plant from the transformed plant cell. Another method of making a transgenic plant comprising the nucleic acid construct of the present invention comprises transforming a seed or immature plant with a vector comprising the ENDO promoter operably linked to a heterologous gene, growing the seed or plant to maturity, obtaining the seeds of the plant, and generating transgenic plants from the seeds. The transgenic plants of the present invention are useful in that they may express a gene product for a desired property such as insect resistance, pesticide resistance, heat, cold or drought tolerance, herbicide tolerance, improved properties, and so on.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] Fig. 1 depicts an RNA gel blot analysis of RC 15 gene expression. Five micrograms of total RNA isolated from stem (St), leaf (L), flower (F), root (R), silique without seed (Si), silique with seeds (Sc) and seedling (Sd) were used to make the RNA gel blot.

[0028] Fig. 2 is a graph comparing relative promoter activity of the 35S and ENDO promoters, as determined by fluorometric beta-glucuronidase (GUS) assays. Light bars represent

GUS activity driven by the ENDO promoter. Dark bars represent GUS activities driven by the 35S promoter. Expression was determined in 1 (rosette), 2 (flower), 3 (stem), 4 (silique), and (30 mature seed).

EXAMPLES

Example 1

[0029] An *Arabidopsis* root cDNA library was constructed using the lambda ZAPII cDNA library construction kit (Stratagene, La Jolla, CA). The cDNA library was converted into plasmid library by mass excision, and bacterial clones were ordered into 384-well microtiter plates. Replica filters were made by gridding four 384-well plates onto a 12cm x 8cm nylon filter using a Biomek 2000 robot. The filter has a 3x3 grid in each well location. As a result, 1,536 cDNA clones were represented on each nylon filter with each clone having its unique duplicate pattern to eliminate false hybridization signals.

[0030] Each replica filter was then hybridized with cDNA probes prepared from various *Arabidopsis* tissues or organs. The random-primed PCR technique (RP-PCR) described by Li et al. (1998) Plant Cell 10:383 was used for cDNA probe preparation to increase the detection sensitivity. Four filters were constructed to represent 6,144 anonymous cDNA clones. Seven filter sets were made and hybridized to RP-PCR probes synthesized from *Arabidopsis* root, leaf, stem, whole silique, silique without seeds, seedling, and flower tissues. The hybridization of each colony to each probe was recorded in a Microsoft Excel database. Sorting algorithms were then utilized to determine the colonies representing putative tissue-specific or constitutive genes. Thirty-three putative constitutive clones were identified and subjected to further characterization. The identity of each clone was determined by DNA sequencing and database searches. RNA dot blot analysis was used to assess the expression pattern of each candidate gene. In order to gain

quantitative information on each clone's expression level, RNA dot blot analysis was performed using RNAs isolated from a transgenic *Arabidopsis* line that harbors a 35S:GUS construct (obtained from the *Arabidopsis* Biological Resource Center (ABRC) at Ohio State University, 1735 Neil Avenue, Columbus, Ohio 43210). Relative transcription activity of each gene was then directly compared with the 35S promoter by incorporating a GUS probe in RNA dot blot analysis.

[0031] A clone designated RC 5 was among the putative constitutive clones chosen for further characterization based on its high levels of expression revealed by RNA dot blot analysis. RNA gel blot analysis showed that CT2 is expressed in most of the tissues examined (stem, leaf, flower, root, silique without seeds, whole silique and seeding) at high levels. (Fig. 1). Sequence analysis indicated that RC15 encoded a putative endomembrane-associated protein. Specifically, a BLASTN search against the GenBank database found that RC15 has been fully sequenced by the EU *Arabidopsis* sequencing project. The BAC clone that contains RC15 is designated F1C12 and was deposited in GenBank as Accession number AL022224 by the EU *Arabidopsis* sequencing project. A portion of FlC12 encodes an endomembrane-associated protein.

[0032] F1C12 was retrieved from GenBank, and the location of the putative RC15 promoter region was identified. Two primers were designed to amplify the putative promoter region (corresponding to bases 69721-71740 of F1C12) upstream of the putative translation start site of RC15. The amplified 2 kb promoter fragment is shown as SEQ ID NO:2. SEQ ID NO:2 depicts the amplified product and contains artificial restriction sites (HindIII site at the 5'-end and BamHI site at the 3'-end) introduced during PCR to facilitate cloning. The amplified promoter was isolated and cloned upstream of the plant reporter gene GUS by replacing the 35S promoter

of the plant transformation vector pBI121 (Bevan (1984) Nucleic Acids Res. 12:8711) with the amplified promoter. The resulting construct is designated pENDO:GUS.

Example 2

[0033] The plasmid pENDO::GUS described in Example 1 was used to transform wild type *Arabidopsis* via vacuum filtration as described by Bechtold et al. (1993) Life Sciences 316:1194. Transgenic plants were recovered and assayed for GUS expression according to Jefferson et al. (1987) EMBO J. 6:3901. Histochemical GUS staining analysis showed that the ENDO promoter drives GUS expression in most of the transgenic plant tissues.

[0034] GUS gene expression was found constitutively in all of the major plant tissues. Further, the ENDO promoter drives GUS gene expression throughout the plant life cycle. Thus the ENDO promoter is a constitutive promoter that drives transgene expression in all of the major plant tissues.

Example 3

[0035] A fluorometric GUS assay was performed to determine the ENDO promoter activity quantitatively according to Jefferson et al., id. Six to eight independent lines were assayed for ENDO promoter activity. Fig. 2 is a graph showing a comparison of relative promoter activity between the 35S and ENDO promoters by fluorometric GUS assays. In Fig. 2, standard deviations are shown above the bars corresponding to activity driven by the ENDO promoter. The y-axis indicates the GUS activity in pmole/mg/min.

[0036] As shown in Fig. 2, the activity of the ENDO promoter is comparable to or greater than the 35S promoter. In developing seed, GUS gene expression is relatively low compared to the 35S promoter. Thus the CT2 promoter is ideal for applications in which strong constitutive expression of a transgene is desired in plant tissues except seed. One example is the expression

of pesticide resistance genes or insecticidal genes for crop protection, in which expression of such genes in seeds is not desirable.